

**PENERAPAN TEXT MINING DAN VECTOR SPACE MODEL
PADA WEB-BASE KNOWLEDGE MANAGEMENT SYSTEM
(STUDI KASUS TEKNIK INFORMATIKA UPN)**

TUGAS AKHIR



Disusun Oleh :

VIVIN SOFI AMALIAH
NPM. 0534010296

**JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN”
JAWA TIMUR
2011**

KATA PENGANTAR

Puji syukur senantiasa penulis ucapkan kehadiran Tuhan Yang Maha Esa, yang telah melimpahkan rahmat dan hidayah-Nya, sehingga penulis dimudahkan dalam penyelesaian penulisan laporan Tugas Akhir ini.

Selama pelaksanaan Tugas Akhir dan dalam penyelesaian penulisan laporan Tugas Akhir ini, penulis mendapatkan banyak bantuan dan bimbingan dari berbagai pihak. Karena itu, penulis ingin mengucapkan terima kasih kepada :

1. Bapak Ir. Sutiyono. MT, selaku dekan Fakultas Teknologi Industri Universitas Pembangunan Nasional "Veteran" Jawa Timur yang telah memberikan motivasi dalam pelaksanaan Tugas Akhir.
2. Bapak Basuki Rahmat. S.si, MT, selaku Kepala Jurusan Teknik Informatika Universitas Pembangunan Nasional "Veteran" Jawa Timur dan dosen pembimbing yang telah memberikan motivasi, saran dan waktu luang dalam menyelesaikan Tugas Akhir.
3. Bapak Moch. Irwan Afandi S.T, M.sc, selaku dosen pembimbing yang telah meluangkan waktu serta sabar dalam memberikan arahan, saran dan motivasi dalam menyelesaikan Tugas Akhir.
4. Orang tua dan keluarga atas segala motivasi dan doanya, sehingga semua dapat berjalan lancar. Dan akhirnya penulis bisa membuat tersenyum keluarga lagi.

5. Mbak is, mas jefry, dik ninit, makasih atas do'a dan motivasinya. Semua keponakan penulis (aril dan ulid) makasih sudah menghibur penulis.
6. Teman – teman yang selalu memberikan motivasi dan semangat dalam pelaksanaan Tugas Akhir khususnya Mohammad Faizin (terima kasih sudah mengajari penulis, semoga ilmu yang kamu sampaikan bermanfaat), Evy Widya Kurniasih (teman seperjuangan), Ainur Rosyidah dan Deffry Hutayare (makasih sudah dipinjami lepi), arin widya, siti hanifah serta teman-teman kos MA1F14 dan MA1A10.
7. Semua anak didik penulis yang selalu menghibur tiap hari. Semoga ilmu yang kalian terima bisa bermanfaat untuk kehidupan nanti.
8. Teman-teman UKKI, dan teman-teman angkatan 2005 serta semua pihak yang tidak dapat disebutkan satu persatu.

Penulis menyadari sepenuhnya masih terdapat banyak kekurangan dalam penyelesaian penulisan laporan Tugas Akhir ini. Namun penulis berusaha menyelesaikan laporan ini dengan sebaik mungkin.

Segala kritik saran yang bersifat membangun sangat diharapkan dari semua pihak, guna perbaikan dan pengembangan dimasa yang akan datang. Akhirnya besar harapan penulis agar laporan ini dapat diterima dan berguna bagi semua pihak. Amin...

Surabaya, Juni 2011

Penulis

DAFTAR ISI

ABSTRAKSI	i
KATA PENGANTAR	ii
DAFTAR ISI	iv
DAFTAR GAMBAR	vii
DAFTAR TABEL	ix
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan Penelitian	3
1.4. Batasan Masalah	3
1.5. Metode Penelitian.....	4
1.6. Sistematika Penulisan.....	5
BAB II LANDASAN TEORI	7
2.1. Teknik Informatika UPN Jatim.....	7
2.2. Knowlegde Management system	8
2.2.1. Alur Pengelolaan Pengetahuan	10
2.2.2. Macam – Macam Bentuk Pengetahuan	11
2.3. Website.....	12
2.4. Ajax.....	14
2.4.1. XMLHttpRequest	16

2.5. JavaScript	17
2.6. JQuery.....	18
2.7. PHP	20
2.8. MySQL	22
2.9. Text Mining	24
2.9.1. Algoritma <i>TF/IDF (Term Frequency – Inversed Document</i> <i>Frequency)</i>	27
2.9.2. Algoritma <i>Vector Space Model</i>	29
2.9.3. Pengkategorian Artikel.....	32
BAB III ANALISIS DAN DESAIN SISTEM	33
3.1. Desain Sistem	33
3.1.1. Perancangan Sistem	34
3.1.2. ER Diagram Sistem	36
3.2. Analisis Sistem	42
3.2.1. Pengolahan data menggunakan algoritma <i>Text Mining – Vector</i> <i>Space Model</i>	42
3.2.2. Forum <i>Realtime</i>	51
BAB IV IMPLEMENTASI	53
4.1. Kebutuhan Sistem	53
4.2. Penerapan Ajax Dalam Membangun Website <i>Knowledge</i> <i>Management System</i>	54

4.3. Pengumpulan Knowledge / Pengatahuan Dengan Menggunakan	
Algoritma Text Mining	58
4.4. Pencarian Pengetahuan Dengan Metode Vector Space Model	
Dari IR (Information Retrieval) Model.....	63
4.5. Transfer Pengetahuan / Knowledge Dengan Memanfaatkan	
Forum	67
BAB V UJI COBA SISTEM DAN EVALUASI.....	70
5.1. Uji Coba Sistem Sistem	70
5.1.1. Halaman Utama.....	71
5.1.2. Halaman Registrasi User Baru Dengan Teknik Ajax	72
5.1.3. Proses Input Artikel Pengatahuan / Knowledge.....	73
5.1.4. Pecarian Artikel Pengetahuan / Knowledge.....	79
5.1.5. Halaman Forum	83
BAB VI PENUTUP	86
6.1 Kesimpulan	86
6.2 Saran	87
DAFTAR PUSTAKA	88

DAFTAR GAMBAR

2.1. Proses Transfer Pengetahuan / Knowledge.....	10
2.2. Perbedaan Model Aplikasi Web Tanpa Dan Dengan Ajax.....	14
2.3. Tahapan Dalam Proses Text Mining	25
2.4. Tahap Tokenizing	25
2.5. Tahap Filtering.....	26
2.6. Tahap Stemming	26
2.7. Tahap Tagging	27
3.1. Alur Pengelolaan Pengetahuan	33
3.2. Sistem KMS Secara Umum	34
3.3. Alur Sistem KMS.....	35
3.4. Proses Transfer Pengetahuan	36
3.5. <i>conceptual data model</i> (CDM).....	37
3.6. <i>physical data model</i> (PDM)	38
3.7. Input Dan Pencarian Pengetahuan	43
3.8. Proses <i>Text Mining</i>	45
3.9. Tabel Untuk <i>Text Mining</i>	47
3.10. Flowchart Proses Analisa Terhadap Input Text	48
3.11. Proses Pencarian Dengan <i>Vector Space Model</i>	49
3.12. Forum Realtime.....	52
4.1. Formulir Dengan Teknik Ajax.....	56
4.2. Validasi Formulir Dengan Ajax	58

4.3. Form Input Artikel Pengatahuan	59
4.4. Form Pencarian	63
4.5. Halaman Forum.....	69
5.1. Halaman Utama Website	70
5.2. Formulir Untuk Pendaftaran User Baru	71
5.3. Validasi Formulir Dengan Data Yang Benar	72
5.4. Validasi Formulir Dengan Data Yang Salah.....	72
5.5. Form Input Artikel Pengetahuan / <i>Knowledge</i>	73
5.6. Hasil Input Artikel Pengetahuan	75
5.7. Hasil Input Artikel Pengetahuan	77
5.8. Hasil Pencarian.....	79
5.9. Halaman Forum.....	85
5.10. Pemberitahuan Pesan Baru Pada Halaman Forum.....	84

DAFTAR TABEL

2.3. Jenis Data PadaMySQL	23
2.3. Perhitungan TD-IDF	28
2.4. Perhitungan <i>Vector-Space Model</i>	29
2.4. Hasil <i>Vector-Space Model</i>	31
3.1 Kategori.....	39
3.2 Artikel	39
3.3 Kata Documen.....	40
3.4 Kata	40
3.5 Stop List.....	40
3.6 Komentar.....	40
3.7 Forum	41
3.8 Thread	41
3.9 Replies.....	41
3.10 Feedback	42
3.12 User	42
5.1. Perbandingan isi artikel terhadap kata kunci kategori dengan perhitungan manual Contoh 1	75
5.2. Perbandingan isi artikel terhadap kata kunci kategori dengan perhitungan manual Contoh 2	77
5.3. Hasil perhitungan dan <i>shorting</i> artikel pengetahuan.....	82



Nama : Vivin Sofi Amaliah
NPM : 0534010296
Judul : Penerapan Text Mining dan Vector Space Model pada Web-Based Knowledge Management System (studi kasus : Teknik Informatika UPN)
Pembimbing 1 : Basuki Rahmad, S.Si, MT
Pembimbing 2 : Moch. Irwan Afandi, ST, M.sc

Abstraksi

Era globalisasi juga ditandai dengan perkembangan ilmu pengetahuan dan teknologi yang sangat pesat. Kemampuan suatu negara di bidang ilmu pengetahuan dan teknologi menjadi salah satu faktor daya saing yang sangat penting dewasa ini. Menyadari akan persaingan global yang semakin ketat dan berat, maka perlu perubahan paradigma dari semula mengandalkan pada *resources-based competitiveness* menjadi *knowledge-based competitiveness*. Maka dalam skripsi ini kami membuat *Knowledge Management System* untuk mengelola pengetahuan dengan menggunakan algoritma *Text Mining* metode *Vector Space Model* dari IR (*Information Retrieval*) Model, agar pengelolaan pengetahuan yang berbentuk text bisa teratur dan terstruktur. Bentuk dari sistem yang kami buat merupakan sebuah web portal yang mempunyai fungsi – fungsi pengelolaan pengetahuan diantaranya transfer pengetahuan dengan cara menginputkan pengetahuan dalam portal tersebut, penyimpanan pengetahuan dalam bentuk digital di database dan pemanfaatan pengetahuan oleh halayak umum. Selain penginputan dan penyimpanan pengetahuan terdapat juga suatu forum yang bisa dimanfaatkan untuk membahas suatu masalah sehingga transfer pengetahuan bisa berjalan dengan efektif. Portal KMS ini diharapkan mampu memfasilitasi tumbuh kembangnya budaya saling berbagi pengetahuan (*share knowledge*) sehingga dapat menciptakan pengetahuan baru yang kompetitif, *decision support system*, sarana penyampaian aspirasi dan penyimpanan dokumen elektronik. Seperti portal, *databases*, *software* dll.

Kata Kunci: *Knowledge Management, Text Mining, Vector Space Model.*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Era globalisasi ditandai dengan perkembangan ilmu pengetahuan dan teknologi yang sangat pesat. Kemampuan suatu negara di bidang ilmu pengetahuan dan teknologi menjadi salah satu faktor daya saing yang paling penting dewasa ini. Sedangkan pengetahuan berkembang lambat dan statis serta belum dapat disinergikan satu dengan yang lainnya.

Seperti halnya di Jurusan teknik informatika UPN Jawa Timur, kebanyakan mahasiswa dan Dosen hanya menguasai bidang tertentu saja. Akibatnya ketergantungan pada satu orang yang menguasai bidang spesifik sangat tinggi sehingga aset pengetahuan sebagai *intangible asset* tidak dapat dieksplorasi untuk meningkatkan kemampuan. Sehubungan dengan hal tersebut, dibutuhkan suatu cara untuk mengeksplorasi aset pengetahuan sebagai *intangible asset* melalui pengelolaan pengetahuan (*Knowledge Management*) dengan memanfaatkan teknologi informasi. Dengan memanfaatkan teknologi website untuk membentuk suatu *Knowledge Management System* yang dapat merencanakan, mengumpulkan dan mengorganisir, memimpin dan mengendalikan data yang telah digabung dengan berbagai bentuk pemikiran dari macam – macam sumber yang kompeten.

Text mining merupakan proses pengambilan data berupa teks dari sebuah sumber dalam hal ini sumbernya adalah dokumen. Dengan *text mining* dapat

dicari kata-kata kunci yang dapat mewakili isi dari suatu dokumen lalu dianalisa dan dilakukan pencocokan antara dokumen dengan *database* kata kunci yang telah dibuat untuk menentukan atau memilah kategori suatu dokumen. Sedangkan proses pengukuran tingkat similaritas antar dokumen dilakukan dengan membandingkan suatu kata kunci dengan dokumen. Kata kunci yang digunakan didapat dari proses *ekstraksi* dokumen pada proses pemilahan kategori dokumen.

Agar hasil pengukuran tingkat similaritas dokumen dengan kata kunci mendapatkan hasil yang optimal maka digunakan algoritma *text mining* dimana dalam prosesnya digunakan algoritma TF-IDF (*Term Frequency – Inversed Document Frequency*) dan VSM (*Vector-Space Model*) dari IR (*Information Retrieval*) model untuk mencari nilai *Cosine* (menghitung nilai cosinus sudut antara dua vector) sebagai pengukur tingkat *similaritas* antara dokumen dengan *keyword* yang didapat dari ekstraksi teks pada dokumen.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, telah diambil permasalahan yang akan dibahas dalam penelitian ini adalah bagaimana membuat suatu sistem yang mampu menangani permasalahan dalam pengelolaan pengetahuan diantaranya sebagai berikut:

- a. Bagaimana cara mengidentifikasi dan mengumpulkan *knowledge*?
- b. Bagaimana terjadi dan terbentuknya jaringan pertukaran *knowledge*?
- c. Bagaimana pengkategorian dan perhitungan similaritas dokumen dengan menggunakan *text mining* dan *vector space model* ?

1.3 Tujuan Penelitian

Adapun tujuan yang diharapkan dalam penelitian ini adalah menyusun suatu sistem aplikasi *Web portal Knowledge Management* yang bermanfaat serta dapat memadukan pengetahuan *tacit* dan *explicit* sehingga menjadi suatu kesatuan *knowledge management* untuk mendukung timbulnya penggunaan, pemanfaatan dan pembuatan pengetahuan di Teknik Informatika UPN Jawa Timur dengan menerapkan *Text Mining* dan *Vector Space Model*.

1.4 Batasan Masalah

Batasan masalah yang diambil pada penulisan skripsi ini diharapkan mampu membatasi pembahasan agar sesuai dengan tujuan penelitian itu sendiri.

Adapun batasan masalah yang diajukan adalah sebagai berikut :

- a. Membangun *knowledge management* system yang berbasis website
- b. Text mining digunakan untuk pengkategorian dan pencarian
- c. Vector Space Model digunakan untuk pengukuran kemiripan dokumen
- d. Sistem ini mengolah pengetahuan yang diinputkan oleh user yang telah menjadi anggota.

1.5 Metode Penelitian

Adapun metode penelitian yang digunakan adalah sebagai berikut:

- a. Studi literatur

Pengumpulan data yang dilakukan dengan mencari bahan-bahan kepustakaan dan referensi dari berbagai sumber sebagai landasan teori

yang ada hubungannya dengan permasalahan yang dijadikan objek penelitian.

b. Analisa Kebutuhan Sistem

Data dan informasi yang telah diperoleh akan dianalisa agar didapatkan kerangka global yang bertujuan untuk mendefinisikan kebutuhan sistem, di mana nantinya akan digunakan sebagai acuan perancangan sistem.

c. Perancangan sistem

Berdasarkan data dan informasi yang telah diperoleh serta analisa kebutuhan untuk membangun sistem ini, akan dibuat rancangan kerangka global yang menggambarkan mekanisme dari sistem yang akan dibuat.

d. Coding

Tahapan ini menerjemahkan hasil perancangan spesifikasi program dari tahapan sebelumnya kedalam baris-baris kode program yang dapat dimengerti oleh komputer.

e. Eksperimen dan Evaluasi

Pada tahap ini, sistem yang telah selesai dibuat akan diuji coba, yaitu pengujian berdasarkan fungsionalitas program, dan akan dilakukan koreksi dan penyempurnaan program jika diperlukan.

1.6 Sistematika Penulisan

Untuk mempermudah dan memahami pembahasan penulisan skripsi ini, maka sistematika penulisan disusun sebagai berikut :

Bab I : Pendahuluan

Berisi Latar Belakang, Rumusan Masalah, Tujuan Penelitian, Pembatasan Permasalahan, Metode Penelitian dan Sistematika Penulisan.

Bab II : Tinjauan Pustaka

Berisi tentang landasan teori mengenai permasalahan yang berhubungan dengan penelitian yang dilakukan.

Bab III : Perancangan dan Analisis Sistem

Dalam bab ini berisi mengenai analisa kebutuhan sistem baik software maupun hardware yang diperlukan untuk membuat kerangka global yang menggambarkan mekanisme dari sistem yang akan dibuat.

Bab IV : Implementasi Sistem

Berisi tentang implementasi dari perancangan sistem yang telah dibuat.

Bab V : Uji Coba dan Evaluasi

Berisi tentang pengujian sistem yang telah dibuat serta evaluasi sistem yang telah dibangun.

Bab VI : Penutup

Merupakan bab terakhir yang memuat intisari dari hasil pembahasan yang berisikan kesimpulan dan saran yang dapat digunakan sebagai pertimbangan untuk pengembangan penulisan selanjutnya

